基于自适应逼近残差的稀疏表示语音降噪方法

周伟力 贺前华* 王亚楼 庞文丰
(华南理工大学电子与信息学院 广州 510640)

摘 要：该文提出一种基于自适应逼近残差的稀疏表示语音降噪方法。在字典学习阶段基于 K 奇异值分解(K-Singular Value Decomposition, K-SVD)算法获得干净语音谱的过完备字典，在稀疏表示阶段基于权重因子调整后的噪声谱和估计的交叉项对逼近残差持续自适应地更新，并采用正交匹配追踪(Orthogonal Matching Pursuit, OMP)方法对干净语音谱进行稀疏重构。最后结合估计的干净语音谱与带噪语音相位，通过傅里叶逆变换获得重构的干净语音。实验结果表明所提方法在不同噪声和信噪比条件下相比标准的谱减法，稀疏表示语音降噪算法和基于自回归隐马尔可夫模型的降噪方法有更好的降噪效果。

关键词：语音降噪；稀疏表示；K 奇异值分解；正交匹配追踪
中图分类号：TN912.3 **文献标识码**：A **文章编号**：1009-5896(2017)02-0309-07
DOI: 10.11999/JEIT160369

Adapted Stopping Residue Error Based Sparse Representation for Speech Denoising

ZHOU Weili HE Qianhua WANG Yalou PANG Wenfeng
(*School of Electronic and Information Engineering*, *South China University of Technology*, *Guangzhou* 510640, *China*)

Abstract: A sparse representation speech denoising method based on adapted stopping residue error is proposed. Firstly, an over complete dictionary of the clean speech power spectrum is learned by the K-Singular Value Decomposition (K-SVD) algorithm. In the sparse representation stage, the stopping residue error is adaptively achieved according to the estimated cross terms and the noise spectrum which is adjusted by a weighted factor, and the Orthogonal Matching Pursuit (OMP) approach is applied to reconstruct the clean speech spectrum from the noisy speech. Finally, the clean speech is re-synthesis via the inverse Fourier transform with the reconstructed speech spectrum and the noisy speech phase. The experiment results show that the proposed method outperforms the standard spectral subtraction, sparse representation based speech denoising algorithm and the AutoRegressive Hidden Markov Model (AR-HMM) based speech denoising method in terms of subjective and objective measure.
Key words: Speech denoising; Sparse representation; K-Singular Value Decomposition (K-SVD); Orthogonal Matching Pursuit (OMP)

1 引言

在实际环境中语音信号往往会受到各种噪声的干扰，语音降噪的目的是从带噪语音中恢复出原始的干净语音，从而改善受损语音的质量和可懂度。语音降噪可应用于多个领域，例如在语音识别系统中，语音降噪算法的引入降低了待识别语音的背景噪声干扰，有助于提高语音识别的准确率[1]；另外，在无参考语音的情况下，语音质量客观评价方法基于语音降噪算法构造“准干净语音”，采用有参考源模型对带噪语音进行客观质量评价，获得了良好的效果[2]。

目前常用的语音降噪方法主要有维纳滤波法(Wiener Filter, WF)[3]，谱减法(Spectrum Subtraction, SS)[4]，基于统计模型方法(model-based)[5]和基于隐马尔可夫模型(Hidden Markov Model, HMM)的语音降噪方法[6]。而谱减算法由于运算量较少并且易于实现，因此常用于语音信号处理领域。然而传统谱减算法存在一些影响降噪性能的因素，如噪声谱估计误差(noise magnitude errors)和交叉项误差(cross-correlation errors)等。目前已有一些工作[7,8]分析了这些因素对信号处理系统性能的影响，但是这些工作主要集中于语音识别的性能

收稿日期：2016-04-18；改回日期：2016-08-25；网络出版：2016-10-21
*通信作者：贺前华 eeqhhe@scut.edu.cn
基金项目：国家自然科学基金(61571192)，广东省公益项目(2015A010103003)
Foundation Items: The National Natural Science Foundation of China (61571192), The Science and Technology Foundation of Guangdong Province (2015A010103003)

分析上，而针对这些因素的补偿方法目前仍有待进一步研究。

近年来，稀疏表示作为信号处理的一种新方法，旨在给定的过完备字典中用尽可能少的原子表示信号的主要信息。由于语音信号在正交基变换中具有近似稀疏性，因此可以通过构造符合语音信号结构的过完备字典，使得字典原子可以线性表达语音信号，从而获得较好的重构精度。语音信号具有稀疏性的特点为稀疏表示方法应用于语音降噪提供了可能性[9]。不同于传统降噪方法通过减少或去除噪声来获得干净语音，基于稀疏表示的语音降噪方法从过完备字典中选取原子表达干净语音信号，从而把干净语音从带噪信号中分离出来，达到剔除噪声的目的。目前发展的算法中，孙林慧等人[10]提出基于数据驱动字典的稀疏表示语音降噪方法。而 Zhao 等人[11]则在频域上采用近似 K-SVD 算法训练纯净语音的过完备字典，采用最小角回归(Least Angle Regression, LARS)方法获得纯净信号谱的稀疏表示。文献[12]基于 K-SVD 算法和带噪语音构建时域信号字典，利用 OMP 方法重构干净语音。Sigg 等人[13]则提出一种基于 generative dictionary 的语音降噪方法，采用语音、噪声的组合字典以及改进的 LARS 算法重构干净语音信号。

稀疏表示降噪方法在信号重构阶段通过限定稀疏编码(如 MP, OMP)的逼近残差，从而选取出有意义的原子，使得重构的信号逼近干净语音而非带噪语音。逼近残差与噪声密切相关，而目前发展的基于稀疏表示的降噪算法主要通过带噪信号的初始段估计噪声谱[11]或者利用话音活动检测(Voice Activity Detection, VAD)方法估计信号非语音段的噪声方差来计算逼近残差[10,12]，并且在逼近残差计算中没有考虑噪声谱估计误差等因素[14]。而现实场景下大多数的噪声信号是非平稳的，仅在信号的无声段估计和更新噪声谱并不足够，非平稳环境下的低信噪比鲁棒 VAD 算法目前仍是研究的热点。另外虽然利用语音和噪声的组合字典可以获得噪声成分的有效估计[13]，但是这类方法需要单独训练噪声字典，而现实环境中噪声类型不可预知，因此噪声字典的离线训练并不适用于实际应用中。基于稀疏表示的语音降噪需要以短时帧为单位从带噪信号中重构干净语音，而由于噪声谱具有时变特性，在话音间隙估计的逼近残差对于话音活动期间可能并不准确。因此如果逼近残差能够根据噪声谱的变化进行持续自适应的更新，那么稀疏表示提取的原子能够更好地表征干净信号，使得重构语音更接近原始纯净信号。为此，本文提出一种自适应逼近残差的语音降噪算法，该算法基于过完备字典和稀疏表示实现噪声消除。逼近残差采用连续估计方式进行更新，同时为了补偿噪声谱估计误差和交叉项误差，提高逼近残差计算准确性，该算法对噪声谱估计值进行自适应调整，并对交叉项误差进行了估计。更新的逼近残差最后应用于干净信号的稀疏重构中。

2 基于稀疏表示的语音降噪数学模型

假定带噪时域信号 $y(n)$ 由干净语音信号 $x(n)$ 和噪声信号 $d(n)$ 组成：

$$y = x + d \tag{1}$$

两边同时作离散傅里叶变换：

$$Y(\omega) = X(\omega) + D(\omega) \tag{2}$$

对 $Y(\omega)$ 求取功率谱：

$$\left|Y(\omega)\right|^2 = \left|X(\omega)\right|^2 + \left|D(\omega)\right|^2 + X(\omega)D^*(\omega) + X^*(\omega)D(\omega) \tag{3}$$

其中，$*$ 表示复数谱的共轭。可以看到，带噪功率谱 $\left|Y(\omega)\right|^2$ 由干净语音功率谱 $\left|X(\omega)\right|^2$，噪声功率谱 $\left|D(\omega)\right|^2$ 以及干净语音与噪声谱之间的交叉项所组成。考虑 $\left|X(\omega)\right|^2 \in R^M$ 在功率谱字典 $\boldsymbol{\psi}_{\text{ps}} \in R^{M\times N}$ 上具有稀疏性，那么干净信号谱的稀疏表示形式为

$$\left|X(\omega)\right|^2 = \boldsymbol{\psi}_{\text{ps}}\boldsymbol{C},\ \ \|\boldsymbol{C}\|_0 \le T \ll N \tag{4}$$

$\boldsymbol{\psi}_{\text{ps}}$ 中含有 N 个列向量，每个列向量被称为原子，并且在稀疏表示中需要事先求出。$\|\bullet\|_0$ 为 l_0 范数，$\boldsymbol{C}$ 是 N 维的稀疏系数向量，并且限定只有 T 个非零的元素。在带噪信号谱中求解干净语音谱的稀疏表示即是在满足稀疏限制的前提下估计干净语音谱的稀疏系数向量 $\widehat{\boldsymbol{C}}$：

$$\widehat{\boldsymbol{C}} = \arg\min\|\boldsymbol{C}\|_0,\ \ \text{s.t.}\ \ \left\|\left|Y(\omega)\right|^2 - \boldsymbol{\psi}_{\text{ps}}\boldsymbol{C}\right\|_2 \le \varepsilon \tag{5}$$

其中，$\|\bullet\|_2$ 为 l_2 范数，ε 是与噪声相关的逼近残差。最后通过式(4)重构干净信号谱。可以看到，逼近残差的准确估计使得选取的原子可以更好地重构原始干净信号。而由式(3)可知，噪声谱和交叉项的有效估计能够提高逼近残差计算的准确性。

3 交叉项估计

传统降噪算法为了计算简便通常假定噪声信号 $d(n)$ 具有零均值，并且与干净信号 $x(n)$ 不相关，那么交叉项 $X(\omega)D^*(\omega)$, $X^*(\omega)D(\omega)$ 简化为零[15]。因此通过以上假设，干净信号功率谱近似估计：

$$\left|X(\omega)\right|^2 = \left|Y(\omega)\right|^2 - \left|D(\omega)\right|^2 \tag{6}$$

而该假设引入交叉项误差为

$$R(\omega)=X(\omega)D^*(\omega) + X^*(\omega)D(\omega) \tag{7}$$

$R(\omega)$ 为

$$
\begin{aligned}
R(\omega) &= \left(1-\frac{|X(\omega)|^2+|D(\omega)|^2}{|Y(\omega)|^2}\right)|Y(\omega)|^2 \\
&= \left(1-\frac{|X(\omega)|^2+|D(\omega)|^2}{|X(\omega)|^2+|D(\omega)|^2+X(\omega)D^*(\omega)+X^*(\omega)D(\omega)}\right)|Y(\omega)|^2 \\
&= \left(1-\frac{|X(\omega)|^2+|D(\omega)|^2}{|X(\omega)|^2+|D(\omega)|^2+2|X(\omega)||D(\omega)|\cos(\theta_{\mathrm{X}}(\omega)-\theta_{\mathrm{D}}(\omega))}\right)|Y(\omega)|^2 \\
&= \left(\frac{2\sqrt{\xi(\omega)}\cos(\theta_{\mathrm{X}}(\omega)-\theta_{\mathrm{D}}(\omega))}{1+\xi(\omega)+2\sqrt{\xi(\omega)}\cos(\theta_{\mathrm{X}}(\omega)-\theta_{\mathrm{D}}(\omega))}\right)|Y(\omega)|^2
\end{aligned} \tag{8}
$$

其中，$\theta_{\mathrm{X}}(\omega)$ 和 $\theta_{\mathrm{D}}(\omega)$ 为干净信号谱和噪声谱的相位，$\xi(\omega)\triangleq|X(\omega)|^2/|D(\omega)|^2$ 为瞬时信噪比。由式(8)可以看到，若 $d(n)$ 和 $x(n)$ 不相关，$\cos(\theta_{\mathrm{X}}(\omega)-\theta_{\mathrm{D}}(\omega))$ 为 0，那么 $R(\omega)$ 为 0；而当 $\xi(\omega)\to 0$ 或者 $\xi(\omega)\to\infty$ 时，亦即当 $\mathrm{SNR}\to\pm\infty$，$R(\omega)\to 0$，因此在极低或者极高信噪比下，假定 $R(\omega)$ 为零也是合理的。然而在其他信噪比条件下，$R(\omega)$ 并非可以忽略，特别在频带信噪比为 0 dB($\xi(\omega)$=1)附近，$R(\omega)$ 达到了较大值。而在实际应用中，语音降噪算法往往需要工作于低信噪比(如信噪比为 0 dB)的环境噪声下。图 1 画出了带噪语音及其交叉项的频谱曲线，其中带噪语音由 0 dB 信噪比的白噪声嵌入获得，语音帧长为 32 ms。可以看到，与带噪语音功率谱相比，至少在低频部分，交叉项并不能够忽略。

基于以上分析，为了提升噪声谱估计的准确性，需要对交叉项进行合理的估计。带噪语音复数谱可以通过幅度与相位表示为极坐标形式：

$$Y(\omega)=|Y(\omega)|\mathrm{e}^{\mathrm{j}\theta_{\mathrm{Y}}(\omega)} \tag{9}$$

其中，$\theta_{\mathrm{Y}}(\omega)$ 为带噪语音谱的相位。而噪声谱同样可以通过其幅度和相位谱表示为 $D(\omega)=|D(\omega)|\mathrm{e}^{\mathrm{j}\theta_{\mathrm{D}}(\omega)}$，而噪声的幅度谱 $|D(\omega)|$ 未知，可以通过噪声估算方法获得的平均幅度谱来代替。在语音信号分析中，由于相位不会对语音可懂度造成影响，噪声相位 $\theta_{\mathrm{D}}(\omega)$ 可以由带噪语音相位 $\theta_{\mathrm{Y}}(\omega)$ 代替[15]。通过对式(2)做以上替换，可以得到纯净信号复数谱的一种估计：

$$\widehat{X}(\omega)=[|Y(\omega)|-\left|\widehat{D}(\omega)\right|]\mathrm{e}^{\mathrm{j}\theta_{\mathrm{Y}}(\omega)} \tag{10}$$

将式(10)代入到式(7)，可以近似获得交叉项：

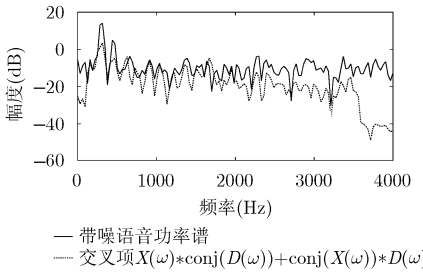

图 1　带噪语音和交叉项频谱曲线，嵌入噪声为 0 dB 白噪声

$$
\begin{aligned}
R(\omega) &= X(\omega)D^*(\omega)+X^*(\omega)D(\omega) \\
&= 2\left(|Y(\omega)|-\left|\widehat{D}(\omega)\right|\right)\left|\widehat{D}(\omega)\right|\left(\mathrm{e}^{\mathrm{j}\theta_{\mathrm{Y}}(\omega)}\mathrm{e}^{-\mathrm{j}\theta_{\mathrm{Y}}(\omega)}\right) \\
&= 2\left(|Y(\omega)|-\left|\widehat{D}(\omega)\right|\right)\left|\widehat{D}(\omega)\right|
\end{aligned} \tag{11}
$$

4　基于自适应逼近残差的稀疏表示语音降噪

为了对逼近残差进行持续更新，通过连续噪声估计方法[16]获得噪声谱估计值 $\left|\widehat{D}(\omega)\right|^2$，并采用与当前帧瞬时后验信噪比相关的权重因子 β [17]进行自适应调整。权重因子主要解决估计噪声谱与瞬时语音谱中实际噪声分量之间可能会存在偏差的问题，通过在低信噪比帧(例如语音的低能量段或没有语音时)对估计的噪声谱施加大的估计权重，而在高信噪比帧(语音成分较大时)施予小的权重，从而达到更好地估计噪声谱的目的。将式(3)表示为第 i 帧带噪信号：

$$|\boldsymbol{Y}_i|^2=|\boldsymbol{X}_i|^2+\beta_i\left|\widehat{\boldsymbol{D}}_i\right|^2+\boldsymbol{R}_i \tag{12}$$

其中，$\boldsymbol{R}_i$ 为第 i 帧的交叉项，由式(11)计算。基于稀疏表示，式(12)的带噪信号功率谱可表示为

$$|\boldsymbol{Y}_i|^2=\boldsymbol{\psi}_{\mathrm{ps}}\boldsymbol{C}_i+\beta_i\left|\widehat{\boldsymbol{D}}_i\right|^2+\boldsymbol{R}_i \tag{13}$$

其中，$\boldsymbol{\psi}_{\mathrm{ps}}$ 为干净信号功率谱的过完备字典，$\boldsymbol{C}_i$ 为第 i 帧的稀疏表示向量，β_i 依据信噪比下设置为

$$\beta_i=\begin{cases}3 & \mathrm{SNR}_i<-5\ \mathrm{dB}\\ 2-\dfrac{1}{10}\mathrm{SNR}_i & -5\ \mathrm{dB}\le\mathrm{SNR}_i\le 10\ \mathrm{dB}\\ 1 & \mathrm{SNR}_i>10\ \mathrm{dB}\end{cases} \tag{14}$$

其中，$\mathrm{SNR}_i(\mathrm{dB})=10\lg\left(\sum_{k=0}^{N-1}|Y_i(k)|^2\Big/\sum_{k=0}^{N-1}\left|\widehat{D}_i(k)\right|^2\right)$，$N$ 为离散傅里叶变换点数。

设置 $\boldsymbol{E}_i=\beta_i\left|\widehat{\boldsymbol{D}}_i\right|^2+\boldsymbol{R}_i$ 的 l_2 范数为第 i 帧带噪信号的自适应逼近残差，干净信号功率谱的稀疏系数向量 $\boldsymbol{C}_i$ 求解为

$$\widehat{\boldsymbol{C}}_i=\arg\min\|\boldsymbol{C}_i\|_0,\quad \mathrm{s.t.}\quad \left\||\boldsymbol{Y}_i|^2-\boldsymbol{\psi}_{\mathrm{ps}}\boldsymbol{C}_i\right\|_2\le\varepsilon_i \tag{15}$$

其中，$\varepsilon_i=\|\boldsymbol{E}_i\|_2=\left\|\beta_i|\widehat{\boldsymbol{D}}_i|^2+\boldsymbol{R}_i\right\|_2$。利用 $\boldsymbol{\psi}_{\mathrm{ps}}$ 和 $\widehat{\boldsymbol{C}}_i$ 对干净信号功率谱进行重构：

$$\left|\widehat{\boldsymbol{X}}_i\right|^2 = \boldsymbol{\psi}_{\mathrm{ps}}\widehat{\boldsymbol{C}}_i \tag{16}$$

基于估计的信号功率谱$|\widehat{\boldsymbol{X}}_i|^2$和噪声谱$\boldsymbol{E}_i$构造带噪幅度谱$|\boldsymbol{Y}_i|$的 MMSE 型噪声抑制滤波器[13]，通过抑制滤波器获得干净信号幅度谱$\left|\widehat{\boldsymbol{X}}_i\right|$：

$$\left|\widehat{\boldsymbol{X}}_i\right| = \boldsymbol{h}_i|\boldsymbol{Y}_i| \tag{17}$$

$\boldsymbol{h}_i$为第i帧带噪信号的噪声抑制滤波器，其表示形式为

$$\boldsymbol{h}_i=\frac{|\boldsymbol{X}_i|}{|\boldsymbol{Y}_i|}=\sqrt{\frac{1-(\gamma_i+1-\boldsymbol{\xi}_i)^2/(4\gamma_i)}{1-(\gamma_i-1-\boldsymbol{\xi}_i)^2/(4\boldsymbol{\xi}_i)}} \tag{18}$$

其中，$\boldsymbol{\xi}_i = \left|\widehat{\boldsymbol{X}}_i\right|^2 \Big/ \boldsymbol{E}_i$，$\gamma_i = |\boldsymbol{Y}_i|^2 \Big/ \boldsymbol{E}_i$分别为瞬时先验和后验信噪比。最后结合$\left|\widehat{\boldsymbol{X}}_i\right|$和带噪语音相位，通过傅里叶逆变换获得重构的干净信号。

式(15)的全局最优解首先需要先求出式(13)中$\boldsymbol{C}_i$的所有解，并从中找出能够满足式(15)限制的最优解。由于字典$\boldsymbol{\psi}_{\mathrm{ps}}$为欠定矩阵(过完备的)，本文采用具有更快收敛速度的 OMP 贪婪算法[18]对$\boldsymbol{C}_i$进行求解。由于语音产生的机理较为复杂，因此基于语音数据训练获得的过完备字典相对传统基函数过完备字典更好地符合语音信号本身的结构[10]。在字典学习方面，K-SVD 算法由于高效，性能优越等特点，目前广泛应用于字典训练中[19]。为此本文基于语音数据，采用 K-SVD 算法构建干净信号功率谱的过完备字典。

本文方法步骤总结如表 1 所示。

5 实验及结果分析

5.1 实验设置

使用 TIMIT 数据库对本文算法进行性能评估，并且采用 NOISEX-92 噪声数据库作为噪声的叠加源。从 TIMIT 数据库训练集中选取 300 段语音，并进行 8k 降采样，帧长取 256 点，帧移 50%，共约 50000 帧样本参与干净语音功率谱字典训练。字典大小为256×1024，字典训练和语音稀疏重构采用 K-SVD 工具箱[20]实现，字典初始化数据从训练样本中随机选取，训练迭代次数为 40。测试样本从 TIMIT 数据库测试集中选取，并使用 White, Babble, F16, Pink 等 4 种不同类型噪声与语音数据合成低信噪比语音样本，信噪比分别为 −5 dB, 0 dB, 5 dB 和 10 dB，共 3200 段样本参与实验评测。将本文方法与文献[4]的标准谱减法(SS)，文献[6]的自回归 HMM 方法(AR-HMM)和文献[11]的频域稀疏表示降噪方法(SRDN)进行比较。其中 AR-HMM 干净语音模型训练数据选自 TIMIT 数据库训练集，持续时长为 20 min，语音 AR 谱阶为 10，状态数为 8，混合态数为 16；而噪声训练数据持续时长为 10 min，每类噪声 HMM 模型 AR 谱阶为 6，状态数为 3，混合态数为 3。通过时域波形和语谱图分析以及客观性能评测两方面验证算法的有效性。

表 1 基于自适应逼近残差的稀疏表示语音降噪

输入：带噪语音y，字典$\boldsymbol{\psi}_{\mathrm{ps}}$。
输出：重构的干净信号$\hat{x}$。
步骤 1　对y进行分帧，获得$y_1, y_2, \cdots, y_M$；
步骤 2　对每帧计算幅度谱$\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_M$和相位$\zeta_1, \zeta_2, \cdots, \zeta_M$；
步骤 3　采用连续噪声估计方法计算噪声功率谱$\left|\widehat{\boldsymbol{D}}_i\right|^2$，$i$=1, 2,⋯,$M$；
步骤 4　计算$\boldsymbol{E}_i=\beta_i\left|\widehat{\boldsymbol{D}}_i\right|^2+\boldsymbol{R}_i$，其中$\boldsymbol{R}_i=2\left|\boldsymbol{Y}_i-\widehat{\boldsymbol{D}}_i\right|\cdot\left|\widehat{\boldsymbol{D}}_i\right|$；
步骤 5　利用式(15)进行 OMP 稀疏分解；
步骤 6　重构每帧干净信号谱：$\left|\widehat{\boldsymbol{X}}_i\right|^2=\boldsymbol{\psi}_{\mathrm{ps}}\widehat{\boldsymbol{C}}_i$；
步骤 7　利用式(17)和式(18)计算$\left|\widehat{\boldsymbol{X}}_i\right|$，结合相位$\zeta_1, \zeta_2, \cdots, \zeta_M$和傅里叶逆变换获得$\hat{x}$。

5.2 时域波形和语谱图分析

图2为原始语音，含噪语音和降噪后的语音时域波形图。其中图2(a)为TIMIT数据库选取的原始语音 (Her wardrobe consists of only skirts and blouses)，图2(b)带噪语音为原始语音叠加10 dB白噪声，图2(c)，图2(d)，图2(e)和图2(f)分别为文献[4]方法、文献[6]方法、文献[11]方法和本文方法重构后的干净语音。图3(a)，图3(b)，图3(c)，图3(d)，图3(e)分别为原始语音，文献[4]方法、文献[6]方法、文献[11]方法和本文方法降噪后语音信号对应的语谱图。

从时域波形可以看到，相对于图 2(c)(文献[4]方法)、图 2(d)(文献[6]方法)和图 2(e)(文献[11]方法)，图 2(f)(本文方法)降噪后的语音更加干净，并且与图 2(a)(原始语音)更为接近。而语谱图方面，图 3 (e)的语音间隙部分有更少的残留噪声，并且相对于图 3(b)，图 3(c)和图 3(d)，图 3 (e)的语音部分更加干净。上述结果表明本文方法相对于比较算法能较好地消除噪声。从时域波形与语谱图发现，相对于原始语音，基于稀疏表示降噪后的语音(图 3(d)，图 3(e))可能会忽略原始语音的某些非语音部分(如句尾的清音‘s’)。其原因可能是清音与白噪声的结构类似，因此在稀疏表示时没有提取表征清音相关的原子，导致重构语音忽略该部分的信息。

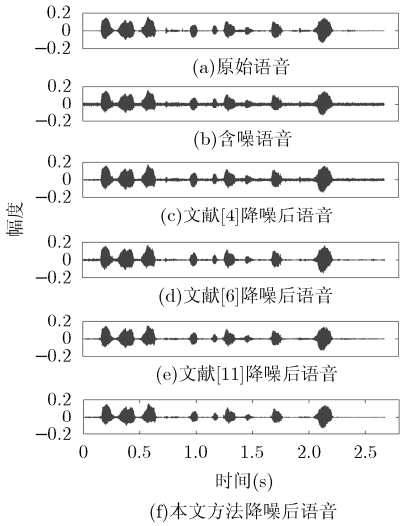

图 2 原始，含噪语音与重构语音波形对比

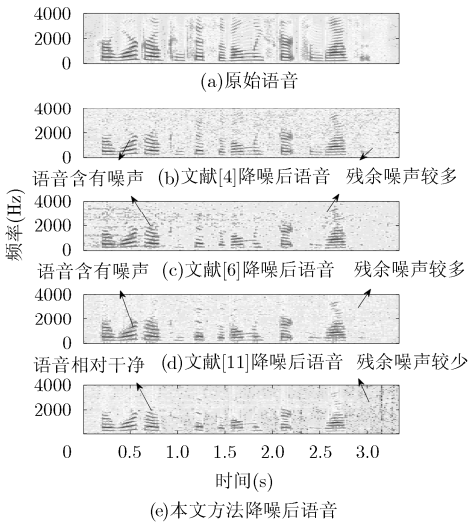

图 3 原始语音与重构语音语谱图

5.3 客观性能评测

采用目前广泛应用的 PESQ 评分[21]和分段信噪比(Segment SNR)客观测度[15]对各种降噪方法进行客观性能评测。图 4 和图 5 为各种降噪算法在不同噪声和信噪比下 PESQ 和 Segment SNR 平均提升幅度的比较结果。Segment SNR 和 PESQ 的提升幅度定义为降噪语音相对干净语音的 Segment SNR 和 PESQ，与原带噪语音相对干净语音的 Segment SNR 和 PESQ 之间的偏差。所有测试样本提升幅度的算术平均作为平均提升幅度。平均提升幅度越大，说明算法的降噪效果越佳。

可以看到，在 PESQ 提升幅度方面，本文方法在−5 dB, 0 dB 和 5 dB 信噪比下，4 种类型噪声相对于对比算法都有更大的提升幅度。而在 10 dB 信噪比下，4 种噪声中有 3 类噪声相对其他比较方法性能更优。在 −5 dB, 0 dB 和 5 dB 信噪比下，本文方法所有噪声的平均提升幅度为 0.31, 0.40 和 0.38。而在 10 dB 信噪比下，所有噪声的平均提升幅度为 0.26。在 Segment SNR 方面，本文方法在−5 dB 和 0 dB 信噪比下，4 种类型噪声相对其他比较方法有

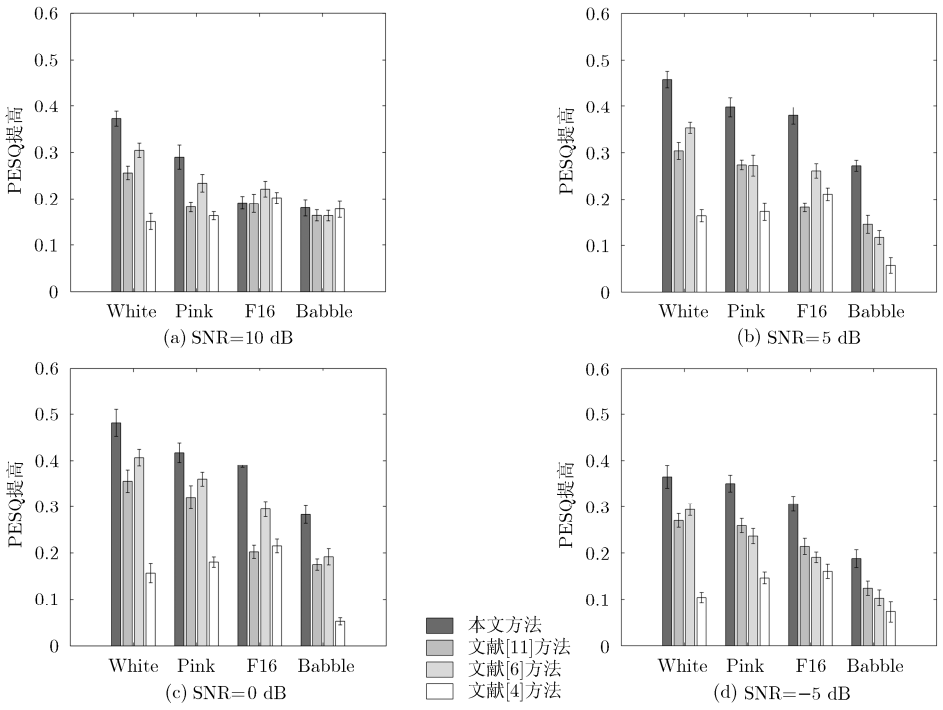

图 4　各种算法 PESQ 平均提升幅度比较 (柱状图代表平均提升的幅度，误差线代表提升幅度 95%的置信区间)

更大的提升幅度。而在 5 dB, 10 dB 信噪比，4 种噪声下有 3 类噪声性能更优。所有噪声在−5 dB, 0 dB 和 5 dB 信噪比下的平均提升幅度为 3.79 dB, 3.18 dB 和 2.02 dB，而在 10 dB 信噪比下的平均提升幅度为 1.26 dB。实验结果表明，本文方法在大部分条件下相对其他比较算法有更好的性能，并且在低信噪比下(−5 dB, 0 dB 和 5 dB)，相对高信噪比(10 dB)性能提升更明显。主要原因可能在于 AR 系数只能模拟语音信号的谱包络，并不能对谱细节成分进行较好的描述，故基于 AR-HMM 降噪算法的语音重构信号在谱细节间仍存在一定的残余噪声；而相对于 SS 和 SRDN 方法，自适应估计的逼近残差使得稀疏表示提取的原子能够更好地表征干净语音，重构后语音更接近原始纯净信号。在低信噪比下，交叉项和权重因子调整后的噪声谱对提高噪声谱估计准确性的作用更大，因此获得的重构语音对带噪语音的改善相对在高信噪比下会更加明显。

6 结束语

本文从信号稀疏重构的角度提出一种自适应逼近残差的稀疏表示语音降噪方法。该方法基于相位不会对语音可懂度造成影响的原则对交叉项进行了近似估计，并通过瞬时后验信噪比相关的权重因子对估计的噪声谱进行调整。在字典训练阶段，基于 K-SVD 算法训练干净语音谱的过完备字典，在稀疏表示时，基于调整后的噪声谱和估计的交叉项自适应地更新逼近残差，并采用 OMP 算法对干净语音谱进行稀疏重构。最后结合重构的干净语音谱和带噪语音相位，通过逆傅里叶变换获得干净语音。在不同噪声和信噪比条件下对重构的干净语音进行主客观评测，实验表明本文方法的有效性。

从实验结果可以看到，算法对于 Babble(多人说话)类型噪声的降噪效果虽然有一定的提高，但是提高幅度并不如其他类型的噪声。有可能 Babble 是一种跟语音相似的结构形背景噪声，其频谱结构与语音有一定的重叠部分，在稀疏表示时提取的原子会表征 Babble 噪声的部分信息，导致重构语音包含部分噪声。因此如果能够在线获得噪声的结构知识(例如在线噪声字典学习)，那么结合这些噪声结构信息可以进一步提高降噪效果，这也是我们下一步的工作。

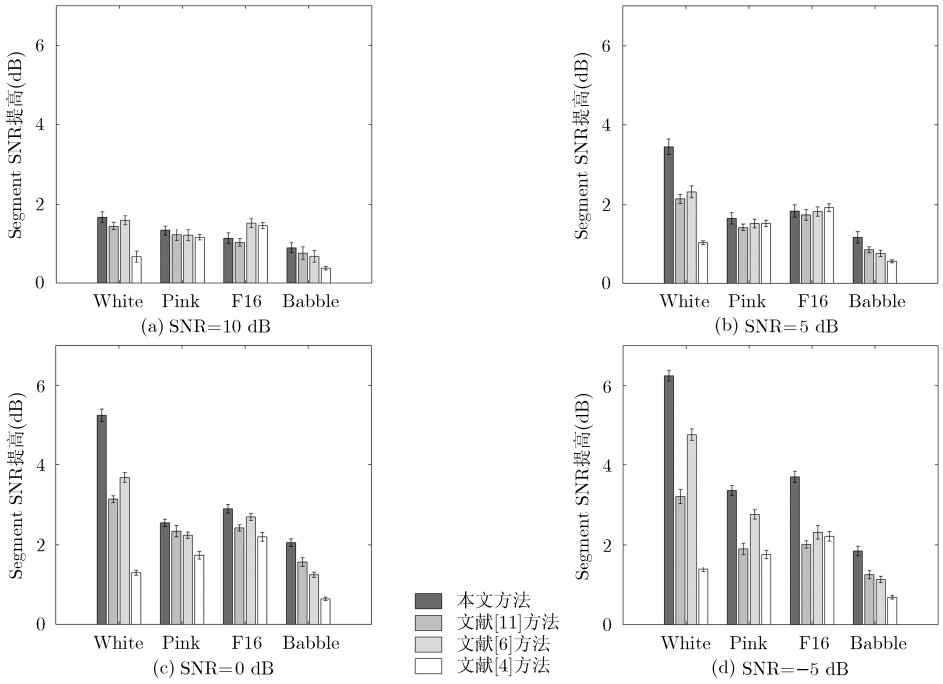

图 5 各种算法 Segment SNR 平均提升幅度比较 (柱状图代表平均提升的幅度，误差线代表提升幅度 95%的置信区间)

参 考 文 献

[1] BABY D, VIRTANEN T, GEMMEKE J F, *et al*. Coupled dictionaries for exemplar-based speech enhancement and automatic speech recognition[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2015, 23(11): 1788–1799. doi: 10.1109/TASLP.2015.2450491.

[2] ZHOU W L and HE Q H. Non-intrusive speech quality objective evaluation in high-noise environments[C]. IEEE China Summit and International Conference on Signal and Information Processing, Chengdu, China, 2015: 50–54. doi: 10.1109/ChinaSIP.2015.7230360.

[3] KODRASI I, MARQUARDT D, and DOCLO S.

Curvature-based optimization of the trade-off parameter in the speech distortion weighted multichannel wiener filter[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, Australia, 2015: 315–319. doi: 10.1109/ICASSP.2015.7177982.

[4] MARTIN R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. *IEEE Transactions on Speech and Language Processing*, 2001, 9(5): 504–512. doi: 10.1109/89.928915.

[5] GERKMANN T. MMSE-optimal enhancement of complex speech coefficients with uncertain prior knowledge of the clean speech phase[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 2014: 4478–4482. doi: 10.1109/ICASSP.2014.6854449.

[6] DAVID Y and KLEIJN W B. HMM-based gain modeling for enhancement of speech in noise[J]. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, 2007, 15(3): 882–892. 10.1109/TASL.2006.885256.

[7] EVANA N, MASON J, LIU W, *et al*. An assessment on the fundamental limitations of spectral subtraction[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Toulous, France, 2006: 145–148. doi: 10.1109/ICASSP.2006.1659978.

[8] HILMAN F, KOJI I, and KOICHI S. Feature normalization based on non-extensive statistics for speech recognition[J]. *Speech Communication*, 2013, 55(5): 587–599. doi: 10.1016/j.specom.2013.02.004.

[9] HSIEH C T, HUANG P Y, CHEN Y H, *et al*. Speech enhancement based on sparse representation under color noisy environment[C]. International Symposium on Intelligent Signal Processing and Communication Systems, Nusa Dua, Indonesia, 2015: 134–138. doi: 10.1109/ISPACS.2015.7432752.

[10] 孙林慧, 杨震. 基于数据驱动字典和稀疏表示的语音增强[J]. 信号处理, 2011, 27(12): 1793–1800.
SUN L H and YANG Z. Speech enhancement based on data·driven dictionary and sparse representation[J]. *Signal Processing*, 2011, 27(12): 1793–1800.

[11] ZHAO Y P, ZHAO X H, and WANG B. A speech enhancement method employing sparse representation of power spectral density[J]. *Journal of Information and Computational Science*, 2013, 10(6): 1705–1714.

[12] ZHAO N, XU X, and YANG Y. Sparse representations for speech enhancement[J]. *Chinese Journal of Electronics*, 2011, 19(2): 268–272.

[13] SIGG C D, DIKK T, and BUHMANN J M. Speech enhancement using generative dictionary learning[J]. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, 2012, 20(6): 1698–1712. doi: 10.1109/TASL.2012.2187194.

[14] ZHAO Y P and WANG B. A speech enhancement method based on sparse reconstruction of power spectral density [J]. *Computers & Electrical Engineering*, 2014, 40(4): 1705–1714. doi: 10.1016/j.compeleceng.2013.12.007.

[15] LOIZOU P C. Speech Enhancement: Theory and Practice [M]. Florida, US: CRC Press, 2013: 104–106.

[16] RANGACHARI S and LOIZOU P. A noise estimation algorithm for highly nonstationary environments[J]. *Speech Communication*, 2006, 48(2): 220–231. doi: 10.1016/j.specom.2006.08.005.

[17] BEROUTI M, SCHWARTZ M, and MAKHOUL J. Enhancement of speech corrupted by acoustic noise[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Washington, US, 1979: 4478–4482. doi: 10.1109/ICASSP.1979.1170788.

[18] CHANG L H and WU J Y. An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit[J]. *IEEE Transactions on Information Theory*, 2014, 60(9): 5702–5715. doi: 10.1109/TIT.2014.2338314.

[19] AHARON M and ELAD M. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322. doi: 10.1109/TSP.2006. Signal 881199.

[20] Ron R. K-SVD ToolBox[OL]. http://www.cs.technion.ac.il/~ronrubin/software.html, 2016.

[21] ITU-T. P.862-2001. Perceptual evaluation of speech quality (PESQ): An objective method for end to end speech quality assessment of narrow-band telephone networks and speech codecs[S]. Geneva, ITU-T, 2001.

周伟力：　男，1986 年生，博士生，从事语音质量客观评价、语音信号降噪的研究工作.

贺前华：　男，1965 年生，博士生导师，教授，研究方向为语音及音频信号处理、嵌入式系统开发.

王亚楼：　男，1991 年生，硕士生，研究方向为音频信号处理.